

# VU Research Portal

## The event-history approach to program evaluation

Abbring, J.H.

2006

### **document version**

Early version, also known as pre-print

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Abbring, J. H. (2006). *The event-history approach to program evaluation*. (TI-Discussion Paper; No. 06-057/3). Tinbergen Instituut (TI).

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)



TI 2006-057/3

Tinbergen Institute Discussion Paper

# The Event–History Approach to Program Evaluation

*Jaap H. Abbring*

*Department of Economics, Vrije Universiteit Amsterdam, and Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# The Event-History Approach to Program Evaluation<sup>\*</sup>

Jaap H. Abbring<sup>†</sup>

August 30, 2006

## Abstract

This paper studies the event-history approach to microeconomic program evaluation. We present a mixed semi-Markov event-history model, discuss its application to program evaluation, and analyze its empirical content. The results of this paper provide fundamental insights in what can be learned from longitudinal micro data about, for example, the effects of training programs for the unemployed on their unemployment durations and subsequent job stability. They can guide the choice of particular models and methods for the empirical analysis of such effects.

---

<sup>\*</sup>Revised version of “The Non-Parametric Identification of Mixed Semi-Markov Event-History Models” (July 2000) for D. Millimet, J. Smith, and E. Vytlacil (2006), editors, *Advances in Econometrics, Volume 21: Modeling and Evaluating Treatment Effects in Econometrics*, Elsevier Science, Oxford.

<sup>†</sup>Department of Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute. Email: jabbring@econ.vu.nl.

Keywords: event-history analysis, identifiability, mixed semi-Markov model, program evaluation  
JEL-codes: C14, C31, C41

This research was supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

# 1 Introduction

Event-history methods are an important tool for the microeconomic evaluation of dynamic programs using longitudinal data. For example, the effects of training and counseling on unemployment durations and job stability have been analyzed by applying event-history methods to data on individual labor-market and training histories (Ridder, 1986; Card and Sullivan, 1988; Gritz, 1993; Ham and LaLonde, 1996; Eberwein et al., 1997; Bonnal et al., 1997). Similarly, the moral hazard effects of unemployment insurance have been studied by analyzing the effects of time-varying benefits on labor-market transitions (*e.g.* Meyer, 1990; Abbring et al., 2005; Van den Berg et al., 2004). In fields like epidemiology, the use of event-history models to analyze treatment effects is widespread (see *e.g.* Andersen et al., 1993; Keiding, 1999).

In this paper, we study the event-history approach to program evaluation.<sup>1</sup> We present a mixed semi-Markov event-history model. We discuss its applications to program evaluation and develop some novel identification results.

The event-history approach to program evaluation is firmly rooted in the econometric literature on state dependence and heterogeneity (Heckman and Borjas, 1980; Heckman, 1981). In the tradition of the selection-model literature, event-history models along the lines of Heckman and Singer (1984a, 1986) are used to jointly model transitions into programs and transitions into outcome states. Causal effects of programs are modeled as the dependence of individual transition rates on the individual history of program participation. Dynamic selection effects are modeled by allowing for dependent unobserved heterogeneity in both the program and outcome transition rates.

Without restrictions on the class of models considered, true state dependence

---

<sup>1</sup>See Heckman and Vytlačil (2006) for a review of the program evaluation literature, Abbring and Heckman (2006) for a study of the microeconomic treatment-effects and structural approaches to dynamic policy evaluation, and Abbring and Van den Berg (2004) for a discussion of the relation between the event-history approach to program evaluation and standard latent-variable and panel-data methods.

and dynamic selection effects cannot be distinguished. Any history dependence of current transition rates can be explained both as true state dependence and as the result of unobserved heterogeneity that simultaneously affects the history and current transitions. This is a dynamic manifestation of the problem of causal inference from observational data. It is the fundamental problem of distinguishing state dependence and heterogeneity.

In applied work, researchers avoid this problem by imposing additional structure. One example is a mixed semi-Markov model in which the causal effects are restricted to the effects of program participation in the previous spell (*e.g.* Bonnal et al., 1997). There is a substantial literature that studies the structure needed to enable the identification of state dependence and heterogeneity in duration and event-history models from longitudinal micro data (see Heckman and Taber, 1994; Van den Berg, 2001, for reviews). However, little is known about the identifiability of general event-history models. The existing literature restricts attention to either single-spell two-state models (*e.g.* Elbers and Ridder, 1982; Heckman and Singer, 1984b; Abbring, 2002), multi-spell two-state models (Honoré, 1993), or competing-risks models (Heckman and Honoré, 1989; Abbring and Van den Berg, 2003a). Neither of these models handle the effect of a dynamically assigned treatment, like a training program, on event-history outcomes such as unemployment durations.

Abbring and Van den Berg (2003b) develop results for a structural bivariate duration model of the effect of a single treatment time on an outcome duration. Their model can be rewritten as a particular three-states event-history model with state dependence (see Section 2). In this paper, we discuss more general event-history models. We will focus on mixed semi-Markov models, which allow for dynamic selection and various forms of state dependence, including duration dependence and dependence on the previous state occupied (“lagged occurrence dependence”).

The paper is organized as follows. Section 2 presents the mixed semi-Markov event-history model and discusses its relation to models that have been used in

empirical work. Section 3 discusses the model’s identifiability from a random sample of censored event histories. Section 4 discusses alternative sampling schemes. Section 5 concludes with some remarks on the implications for applied empirical work.

## 2 A Mixed Semi-Markov Event-History Model

### 2.1 Model

The model is set up along the lines of Heckman and Singer (1984a, 1986). Point of departure is a continuous-time stochastic process assuming values in a finite set  $\mathcal{S}$  at each point in time. We will interpret realizations of this process as agents’ event histories of transitions between states in the state space  $\mathcal{S}$ .

Suppose that event histories start at real-valued random times  $T_0$  in a  $\mathcal{S}$ -valued random state  $S_0$ , and that subsequent transitions occur at random times  $T_1, T_2, \dots$  such that  $T_0 < T_1 < T_2 < \dots$ . Let  $S_l$  be the random destination state of the transition at  $T_l$ . Taking the sample paths of the event-history process to be right-continuous, we have that  $S_l$  is the state occupied in the interval  $[T_l, T_{l+1})$ .

Suppose that heterogeneity between agents is captured by vectors of time-constant observed covariates  $X$  and unobserved covariates  $V$ .<sup>2</sup> Then, state dependence in the event-history process for given individual characteristics  $X, V$  has a causal interpretation.<sup>3</sup> We structure such state dependence by assuming that the event-history process conditional on  $X, V$  is a time-homogeneous semi-Markov process: Conditional on  $X, V$  the origin-destination-specific transition intensities depend only on

---

<sup>2</sup>We restrict attention to time-invariant observed covariates for expositional convenience. The analysis can easily be adapted to more general time-varying external covariates. Restricting attention to time-constant regressors is a worst-case scenario for identification: External time variation in observed covariates aids identification (Heckman and Taber, 1994).

<sup>3</sup>Abbring and Van den Berg (2003b) make their model’s causal structure explicit in a potential-outcomes model of the causal effects of a treatment time on an outcome duration. Abbring (2003) and Abbring and Heckman (2006) present the symmetric extension of this model, a non-parametric structural bivariate duration model allowing for simultaneous causal dependence of both durations. Extending this further to the general event-history setup adds a lot of complexity, but little extra insight.

the current state and the elapsed duration in the current state. In our notation,  $(\Delta T_l, S_l) \perp\!\!\!\perp \{(T_i, S_i), i = 0, \dots, l-1\} | S_{l-1}, X, V$ , where  $\Delta T_l := T_l - T_{l-1}$  is the length of spell  $l$ . Also, the distribution of  $(\Delta T_l, S_l) | S_{l-1}, X, V$  does not depend on  $l$ . Note that, conditional on  $X, V$ ,  $\{S_l, l \geq 0\}$  is a time-homogeneous Markov chain under these assumptions.

Non-trivial dynamic selection effects arise because  $V$  is not observed. The event-history process conditional on observed covariates  $X$  only is a mixed semi-Markov process. If  $V$  affects the initial state  $S_0$ , or transitions from there, subpopulations of agents in different states at some time  $t$  typically have different distributions of the unobserved characteristics  $V$ . Therefore, a comparison of the subsequent transitions in two such subpopulations does not only reflect state dependence, but also sorting of agents with different unobserved characteristics into the different states they occupy at time  $t$ .

We model  $\{(\Delta T_l, S_l), l \geq 1\} | T_0, S_0, X, V$  as a repeated competing-risks model. Due to the mixed semi-Markov assumption, the latent durations corresponding to transitions into the possible destination states in the  $l$ -th spell only depend on the past through the current state  $S_{l-1}$ , conditional on  $X, V$ . This implies that we can fully specify the repeated competing-risks model by specifying a set of origin-destination-specific latent durations, with corresponding transition rates. Let  $T_{jk}^l$  denote the latent duration corresponding to the transition from state  $j$  to state  $k$  in spell  $l$ . We explicitly allow for the possibility that transitions between certain (ordered) pairs of states may be impossible. To this end, define the correspondence  $\mathcal{Z} : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$  assigning to each  $s \in \mathcal{S}$  the set of all destination states to which transitions are made from  $s$  with positive probability.<sup>4</sup> Here,  $\mathcal{P}(\mathcal{S})$  is the set of all subsets of  $\mathcal{S}$  (the “power set” of  $\mathcal{S}$ ). Then, the length of spell  $l$  is given by  $\Delta T_l = \min_{s \in \mathcal{Z}(S_{l-1})} T_{S_{l-1}s}^l$ , and the destination state by  $S_l = \arg \min_{s \in \mathcal{Z}(S_{l-1})} T_{S_{l-1}s}^l$ .

We take the latent durations to be mutually independent, jointly independent

---

<sup>4</sup>Throughout the paper, we assume that  $\mathcal{Z}$  is known. It is important to note, however, that  $\mathcal{Z}$  can actually be identified trivially in all cases considered.



from  $T_0, S_0$ , and identically distributed across spells  $l$ , all conditional on  $X, V$ . This reflects both the mixed semi-Markov assumption and the additional assumption that all dependence between the latent durations corresponding to the competing risks in a given spell  $l$  is captured by the observed regressors  $X$  and the unobservables  $V$ . This is a standard assumption in econometric duration analysis, which, with the semi-Markov assumption, allows us to characterize the distribution of  $\{(\Delta T_l, S_l), l \geq 1\} | T_0, S_0, X, V$  by specifying origin-destination-specific hazards  $\theta_{jk}(t | X, V)$  for the marginal distributions of  $T_{jk}^l | X, V$ .

We assume that the hazards  $\theta_{jk}(t | X, V)$  are of the mixed proportional hazard (MPH) type:<sup>5</sup>

$$\theta_{jk}(t | X, V) = \begin{cases} \lambda_{jk}(t) \phi_{jk}(X) V_{jk} & \text{if } k \in \mathcal{Z}(j) \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

The baseline hazards  $\lambda_{jk} : \mathbb{R}_+ \rightarrow (0, \infty)$  capture duration dependence of the individual transition rates. They have integrals  $\Lambda_{jk}(t) := \int_0^t \lambda_{jk}(\tau) d\tau < \infty$  for all  $t \in \mathbb{R}_+ := [0, \infty)$ . The regressor functions  $\phi_{jk} : \mathcal{X} \rightarrow (0, \infty)$  are assumed to be continuous, with  $\mathcal{X} \subset \mathbb{R}^q$  the support of  $X$ . In applications, these functions are frequently specified as  $\phi_{jk}(x) = \exp(x' \beta_{jk})$  for some parameter vector  $\beta_{jk}$ . We will not make such parametric assumptions. Note that the fact that all regressor functions are defined on the same domain  $\mathcal{X}$  is not restrictive, because each function  $\phi_{jk}$  can “select” certain elements of  $X$  by being trivial functions of the other elements. In the parametric example, the vector  $\beta_{jk}$  would only have nonzero elements for those regressors that matter to the transition from  $j$  to  $k$ . Finally, the  $(0, \infty)$ -valued random variable  $V_{jk}$  is the scalar component of  $V$  that affects the transition from state  $j$  to state  $k$ . Note that we allow for general dependence between the components of  $V$ . This way, we can capture, for example, that agents with lower re-employment

---

<sup>5</sup>The MPH model is an extension of the Cox (1972) proportional hazard model by Lancaster (1979) and Vaupel et al. (1979).

rates have higher training enrolment rates. We normalize

$$\Lambda_{jk}(t^*) = 1 \quad \text{and} \quad \phi_{jk}(x^*) = 1, \quad j \in \mathcal{S}, \quad k \in \mathcal{Z}(j),$$

for some *a priori* chosen  $t^* \in (0, \infty)$  and  $x^* \in \mathcal{X}$ . These normalizations are innocuous because  $V_{jk}$  can capture the scale of  $\theta_{jk}$ .

This fully characterizes the distribution of the transitions  $\{(\Delta T_l, S_l), l \geq 1\}$  conditional on the initial conditions  $T_0, S_0$  and the agents' characteristics  $X, V$ . A complete model of the event histories  $\{(T_l, S_l), l \geq 0\}$  conditional on  $X, V$  would in addition require a specification of the initial conditions  $T_0, S_0$  for given  $X, V$ . It is important to stress here that  $T_0, S_0$  are the initial conditions of the event-history process itself, and should not be confused with the initial conditions in a particular sample (which we will discuss in Section 4). In empirical practice, interest in the dependence between start times  $T_0$  and characteristics  $X, V$  is often limited to the observation that the distribution of agents' characteristics may vary over cohorts indexed by  $T_0$ . The choice of initial state  $S_0$  may in general be of some interest, but is often trivial. For example, we could model labor-market histories from the calendar time  $T_0$  at which agents turn 15 onwards. In an economy with perfect compliance to a mandatory schooling age over 15, the initial state  $S_0$  would be “(mandatory) schooling” for all. Therefore, we will not consider a model of the event history's initial conditions, but instead focus on the conditional model of subsequent transition histories.

Because of the semi-Markov assumption, the distribution of  $\{(\Delta T_l, S_l), l \geq 1\} | T_0, S_0, X, V$  only depends on  $S_0$ , and not  $T_0$ . Thus,  $T_0$  only affects observed event histories through cohort effects on the distribution of unobserved characteristics  $V$ . The initial state  $S_0$ , on the other hand, may both have causal effects on subsequent transitions and be informative on the distribution of  $V$ . For expositional clarity, we will assume that the distribution of unobserved covariates does not vary over

cohorts, or more precisely that  $V \perp\!\!\!\perp T_0 | S_0, X$ , throughout the paper.<sup>6</sup>

An econometric model for transition histories conditional on the observed covariates  $X$  can be derived from the model of  $\{(\Delta T_l, S_l), l \geq 1\} | S_0, X, V$  by aggregating over  $V$ . The exact way this should be done depends on the sampling scheme used. First, in Section 3, we consider sampling from the population of event-histories. We assume that we observe the covariates  $X$ , the initial state  $S_0$ , and the first  $L$  transitions from there. Then, we can model these transitions for given  $S_0, X$  by integrating the conditional model over the distribution of  $V | S_0, X$ .

Next, in Section 4, we briefly discuss more complex, and arguably more realistic, sampling schemes. For example, when studying labor-market histories we may randomly sample from the stock of unemployed at a particular point in time. Because the unobserved factor  $V$  affects the probability of being unemployed at the sampling date, the distribution of  $V | X$  in the stock sample does not equal its population distribution. Moreover, in this case we typically do not observe an agent's entire labor-market history from  $T_0$  onwards. Instead, we may have data on the time spent in unemployment at the sampling date and on labor-market transitions for some period after the sampling date. This complicates matters further.

## 2.2 Applications to Program Evaluation

A number of empirical papers study the effect of a single treatment on some outcome duration or set of transitions. Two approaches can be distinguished. In the first approach, the outcome and treatment processes are explicitly and separately specified. The second approach distinguishes treatment as a separate state in a single event-history model with state dependence.

The first approach is used in a variety of papers in labor economics. Eberwein et al. (1997) specify a model for labor market transitions in which the transition intensities between various labor market states (not including treatment) depend on

---

<sup>6</sup>This can easily be relaxed, but at the expense of some extra notation and technical conditions.

whether someone has been assigned to a training program in the past or not. Abbring et al. (2005) and Van den Berg et al. (2004) specify a model for re-employment durations in which the re-employment hazard depends on whether a punitive benefits reduction has been imposed in the past. Similarly, Van den Berg et al. (2002) analyze the duration up to transition into medical trainee positions and the effect of an intermediate transition into a medical assistant position (a “stepping-stone job”) on this duration.

These models fit Abbring and Van den Berg’s (2003b) framework, or a multi-state extension thereof. The model considered by Abbring and Van den Berg is a bivariate duration model in which realization of the outcome duration censors the treatment duration, and realization of the treatment duration changes the hazard of the outcome durations from that time onwards. We can rephrase this type of model in terms of a simple event-history model with state dependence as follows. Distinguish three states, untreated ( $O$ ), treated ( $P$ ), and the exit state of interest ( $E$ ), so that  $\mathcal{S} = \{O, P, E\}$ . All subjects start in  $O$ , so that  $S_0 = O$ . Obviously, we do not want to allow for all possible transitions between these three states. Instead, we restrict the correspondence  $\mathcal{Z}$  representing the possible transitions as follows:

$$\mathcal{Z}(s) = \begin{cases} \{P, E\} & s = O \\ \{E\} & \text{if } s = P \\ \emptyset & s = E \end{cases}$$

Simple state dependence of the transition rates into  $E$  will already capture a treatment effect in the sense of Abbring and Van den Berg. Not all models in their paper are however included in the simple semi-Markov setup discussed here. In particular, in this paper we do not allow the transition rate from  $P$  to  $E$  to depend on the duration spent in  $O$ . This extension with “lagged duration dependence” (Heckman and Borjas, 1980) would be required to capture one variant of their model.

The model for transitions from “untreated” ( $O$ ) is a competing risks model,

with program enrolment (transition to  $P$ ) and employment ( $E$ ) competing to end the untreated spell. If the unobservable factor  $V_{OE}$  that determines transitions to employment and the unobservable factor  $V_{OP}$  affecting program enrolment are dependent, then program enrolment is selective in the sense that the distribution of  $V_{OE}$ — and then typically that of  $V_{PE}$ — among those who enrol at a given point in time does not equal its distribution among survivors in  $O$  up to that time.<sup>7</sup>

The second approach is used by *e.g.* Gritz (1993) and Bonnal et al. (1997). Consider the following simplified setup. Suppose workers are either employed ( $E$ ), unemployed ( $O$ ), or engaged in a training program ( $P$ ). We can now specify a transition process among these three labor market states in which a causal effect of training on unemployment and employment durations is modeled as dependence of the various transition rates on the past occurrence of a training program in the labor-market history. Partly to avoid initial conditions problems, Bonnal et al. restrict attention to first-order lagged occurrence dependence. So, suppose that transition rates only depend on the current and previous state occupied. Such a model is not directly covered by the semi-Markov model, but with a simple augmentation of the state space it will be. In particular, we have to include lagged states in the state space on which the transition process is defined. Because there is no lagged state in the event-history's first spell, initial states should be defined separately. So, instead of just distinguishing states in  $\mathcal{S}^* = \{E, O, P\}$ , we distinguish augmented states in  $\mathcal{S} = \{(s, s') \in (\mathcal{S}^* \cup I) \times \mathcal{S}^* : s \neq s'\}$ . Then,  $(I, s)$ ,  $s \in \mathcal{S}^*$ , denote the initial states, and  $(s, s') \in \mathcal{S}$  the augmented state of an agent who is currently in  $s'$  and came from  $s \neq s'$ . In order to preserve the interpretation of the model as a model of lagged occurrence dependence, we have to exclude certain transitions by specifying

$$\mathcal{Z}(s, s') = \{(s', s''), s'' \in \mathcal{S}^* / \{s'\}\}.$$

---

<sup>7</sup>Note that, in addition, the survivors in  $O$  themselves are a selected subpopulation: Because  $V$  affects survival in  $O$ , the distribution of  $V$  among survivors in  $O$  is not equal to its population distribution.

This excludes transitions to augmented states that are labeled with a lagged state different from the origin state. Also, it ensures that agents never return to an initial state. For example, from the augmented state  $(O, P)$ — previously unemployed and currently enrolled in a program— only transitions to augmented states  $(P, s'')$ — previously enrolled in a program and currently in  $s''$ — are possible. Moreover, it is not possible to be currently employed and transiting to initially unemployed,  $(I, O)$ . Rather, an employed who loses his job would transit to  $(E, O)$ — currently unemployed and previously employed.

The effects of, for example, training are now modeled as simple state-dependence effects. For example, the effect of training on the transition rate from unemployment to employment is simply the contrast between the individual transition rate from  $(E, O)$  to  $(O, E)$  and the transition rate from  $(P, O)$  to  $(O, E)$ . Dynamic selection into the augmented states  $(E, O)$  and  $(P, O)$ , as specified by the transition model, confounds the empirical analysis of these training effects. Note that there are no longer run effects of training on transition rates from unemployment to employment due to the fact that we have restricted attention to first-order lagged occurrence dependence, like Bonnal et al. (1997).

## 3 The Model's Empirical Content

### 3.1 Sampling Scheme

Suppose that we randomly sample from the population of event histories, and that we observe the first  $L$  transitions, including destinations, for each sampled event-history, with possibly  $L = \infty$ . Thus, we observe a random sample of  $\{(T_l, S_l), l \in \{0, 1, \dots, L\}\}$ , and  $X$ . If  $L < \infty$  then our data are right-censored; if  $L = \infty$  they are not.<sup>8</sup>

---

<sup>8</sup>Our results can be adapted to other common censoring schemes, such as censoring at some non-random finite time. See Andersen et al. (1993) for an overview of censoring schemes.

## 3.2 Identification from First Transitions and Variation in Initial Conditions

First, consider what can be learned from data on the first transition from the initial state  $S_0$  only. Denote the support of  $S_0$  by  $\mathcal{S}_0$ . For  $j \in \mathcal{S}$ , let  $\#\mathcal{Z}(j)$  denote the number of elements in  $\mathcal{Z}(j)$ , i.e., the number of destination states that are reached with positive probability from  $j$ . Consider the following assumptions.

**Assumption 1.**  $(V_{jk}, k \in \mathcal{Z}(j)) \perp\!\!\!\perp X | S_0$ , for all  $j \in \mathcal{S}_0$ .

**Assumption 2.**  $\mathbb{E}[V_{jk}] < \infty$  for all  $k \in \mathcal{Z}(j)$  and  $j \in \mathcal{S}_0$ .

**Assumption 3.** The range  $\{(\phi_{jk}(x); k \in \mathcal{Z}(j)), x \in \mathcal{X}\}$  of the regressor functions contains a nonempty open set in  $(0, \infty)^{\#\mathcal{Z}(j)}$ , for all  $j \in \mathcal{S}_0$ .

These are multivariate extensions of assumptions that are standard in the single-spell MPH literature (e.g. Elbers and Ridder, 1982). Assumption 1 requires independence of the observed covariates and the unobserved heterogeneity in the relevant subpopulations. Because we only observe a single transition from each origin state and cannot apply panel-data techniques to deal with unobserved heterogeneity, this is necessary for the identification of the regressor functions  $\phi_{jk}$ .

Assumption 2 is a technical, but far from innocuous assumption (Ridder, 1990). Without it, the integrated baseline hazards  $\Lambda_{jk}$  and regressor functions  $\phi_{jk}$  can only be identified up to power transformations.<sup>9</sup> Such transformations may substantially change the interpretation of  $\Lambda_{jk}$  and  $\phi_{jk}$ .

Finally, Assumption 3 ensures that there is independent variation with the regressors of the individual hazard rates  $\theta_{jk}(t|X, V)$  corresponding to the various competing risks in state  $j$ . With  $\phi_{jk}(x) = \exp(x'\beta_{jk})$ , it would be sufficient that  $(\beta_{jk}, k \in \mathcal{Z}(j))$  has full column rank and  $\mathcal{X}$  contains a non-empty open set in  $\mathbb{R}^q$ , for all  $j \in \mathcal{S}_0$ . In turn, this could be achieved by imposing exclusion restrictions

---

<sup>9</sup>Ridder and Woutersen (2003) prove semi-parametric identification of a single-spell MPH model under an alternative assumption on the baseline hazard that is equally substantial.

of the sort encountered in instrumental-variables analysis. However, such exclusion restrictions are not necessary for Assumption 3 to hold.

We have the following result.

**Proposition 1.** *If Assumptions 1–3 are satisfied, then  $((\phi_{jk}, \Lambda_{jk}), j \in \mathcal{S}_0, k \in \mathcal{Z}(j))$  and the joint distributions of  $(V_{jk}, k \in \mathcal{Z}(j))|S_0 = j, j \in \mathcal{S}_0$ , are identified from the distribution of  $\Delta T_1, S_1|S_0, X$ .*

*Proof.* For each  $j \in \mathcal{S}_0$ , the model of  $\Delta T_1, S_1|S_0 = j, X$  is an MPH competing-risks model. The result follows from repeated application of Abbring and Van den Berg (2003a, Proposition 2).<sup>10</sup>  $\square$

Because the model for the first transition from the initial state is an MPH competing-risks model, Proposition 1’s proof is a direct application of Abbring and Van den Berg’s (2003a) identification results for such models. The intuition for these results comes in two stages.

First, consider the transition rate from state  $j$  to state  $k$  among those who have survived for some time  $t$  in their initial state  $j$ ,

$$\lambda_{jk}(t)\phi_{jk}(X)\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X].$$

This “crude” hazard rate can be computed from the distribution of  $\Delta T_1, S_1|S_0, X$  that Proposition 1 takes as data. For  $t > 0$ ,  $\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X]$  typically depends on  $X$  and variation of the crude hazard rate with the covariates reflects both these selection effects and the agent-level effects through  $\phi_{jk}(X)$ . However, because of Assumptions 1 and 2,  $\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X]$  reduces to  $\mathbb{E}[V_{jk}|S_0 = j] < \infty$  as  $t \downarrow 0$ . Thus, near the start of the spell, by Assumption 1 subpopulations with different regressor values are similar in terms of their unobserved components. Therefore,

---

<sup>10</sup>Abbring and Van den Berg study the case with two risks, but the extension to more than two risks is trivial.



we can identify  $\phi_{jk}$  by contrasting crude hazard rates between such subpopulations near the start of the spell.

Second, note that, for  $t > 0$  and given  $\phi_{jk}(X)$ , the crude hazard rate above can only depend on  $X$  through the selection effects on

$$\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X] = \mathbb{E}[V_{jk}|T_{jk}^1 \geq t, k' \in \mathcal{Z}(j); S_0 = j, X].$$

Now suppose that  $V_{jk} \perp\!\!\!\perp (V_{jk'}, k' \in \mathcal{Z}(j)/\{k\})|S_0, X$ . Then, the event  $\{T_{jk'}^1 \geq t, k' \in \mathcal{Z}(j)/\{k\}\}$  is not informative on  $V_{jk}$  for given  $S_0, X$ , and

$$\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X] = \mathbb{E}[V_{jk}|T_{jk}^1 \geq t, S_0 = j, X].$$

Thus, in this case,  $\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X]$ , and therefore the crude hazard rate, does not depend on  $X$  for given  $\phi_{jk}(X)$ . If  $V_{jk} \not\perp\!\!\!\perp (V_{jk'}, k' \in \mathcal{Z}(j)/\{k\})|S_0, X$ , on the other hand,  $\mathbb{E}[V_{jk}|\Delta T_1 \geq t, S_0 = j, X]$  depends on  $\phi_{jk'}(X)$  for given  $\phi_{jk}(X)$  through the dependence of  $V_{jk}$  and  $V_{jk'}$ . In sum, independent variation in  $\phi_{jk'}(X)$ ,  $k' \in \mathcal{Z}(j)/\{k\}$ , and  $\phi_{jk}(X)$  can be exploited to infer whether the competing risks are dependent or not. Assumption 3 ensures that there is such independent variation in the regressor effects.

In applications, like those in Section 2, we are typically interested in contrasting the distributions of  $T_{jk}^1$  and  $T_{j'k}^1$  in a subpopulation with given values of  $X$  and  $S_0$ , for some  $j, j' \in \mathcal{S}_0$  and  $k \in \mathcal{Z}(j) \cap \mathcal{Z}(j')$ . Such a contrast can be interpreted as an average treatment effect on the given subpopulation.<sup>11</sup> However, Proposition 1 only provides identification conditional on  $S_0$ . In particular, it gives conditions under which we can construct the distributions of  $T_{jk}^1|S_0 = j, X$  and  $T_{j'k}^1|S_0 = j', X$ .

---

<sup>11</sup>Abbring and Van den Berg (2005) discuss the definition of treatment effects in duration models. They argue that the usual treatment effects defined in terms of the distributions of potential outcome durations confound effects on individual hazard rates and effects that operate through dynamic selection. Recursive economic models often primarily predict effects on individual hazard rates and semi-parametric structure, such as the MPH model, is needed to identify such effects. Here, we do not explicitly address this issue.

The contrast between these distributions reflects both causal treatment effects and selection into the initial state. This is the standard problem of causal inference. Some standard solutions to this problem, adapted to this event-history setting, are the following.<sup>12</sup>

First, we could assume that assignment to initial states is “randomized”:

**Assumption 4.**  $V \perp\!\!\!\perp S_0$ .

This, with Assumption 1, would allow us to identify the distributions of  $T_{jk}^1|X$  and  $T_{jk}^1|X, S_0$ —our objects of interest—with that of  $T_{jk}^1|S_0 = j, X$ . The latter is identified by Proposition 1.

Second, we could rely on instruments to generate random variation in  $S_0$ . Abbring and Van den Berg’s (2005) non-parametric and semi-parametric results for single-spell duration outcomes apply directly to the special case that there is only one destination from the initial states. Their extension to competing-risks outcomes is required in the general case. The MPH structure will prove key here in separating the effects on the competing risks.<sup>13</sup>

### 3.3 Dynamic Selection

The event-history approach to program evaluation does not rely on random variation in the initial state, but instead exploits variation in the states that arises in the course of the event history due to transitions. There is a close connection to the selection-model literature: Dynamic selection into states is modeled jointly with outcomes by means of the mixed semi-Markov event-history model.

This approach is particularly relevant in the important case in which there is no variation at all in the initial state, and  $S_0$  is degenerate. Then, Assumption 4 is trivially satisfied, but there is no scope for contrasting transitions from different

---

<sup>12</sup>See *e.g.* Heckman et al. (1999) for an overview.

<sup>13</sup>In addition, semi-parametric structure will be important if one is interested in treatment effects on individual hazard rates. See Footnote 11.

initial states. In this case, all variation in states arises dynamically according to the mixed semi-Markov transition process. Multiple spells ( $L > 1$ ) are needed to compare transition rates from different origin states.

A selection problem arises if the probability, conditional on  $X, V$ , that the state of interest is never occupied during the observed event history depends on  $V$  for given  $X$ . This is typically the case if only a finite number  $L$  of spells is observed, so that there is censoring.<sup>14</sup> Without censoring, if  $L = \infty$ , it is often true that the state of interest is almost surely occupied at some point during the event history. Then, the sample of event histories' first spells in the state of interest does not suffer from selection on unobservables; Standard competing-risks results then give full identification. Similarly, if both the first and the second spell in a state of interest occur almost surely, the much stronger results for multi-spell competing-risks models of Abbring and Van den Berg (2003a) can be applied.

Obviously, these results are of little empirical relevance, as we typically only observe a limited number of spells in any event-history, and panel data are subject to other types of censoring. However, they highlight that dynamic selection problems arise either because of restrictions on the event-history process that lead to selective occurrence of first (and higher) spells in given states, or because of limited observability of event histories, *e.g.* due to censoring.

In the remainder of this section we will analyze under what conditions the structure imposed on the mixed semi-Markov model in Section 2 suffices for identification of state dependence and treatment effects.

### 3.4 Identification from Censored Event Histories

Let  $j \in \mathcal{S}$  be accessible from  $S_0$ , so that  $\Pr(S_l = j) > 0$  for some  $l \geq 0$ , and let  $k \in \mathcal{Z}(j)$ . Consider the identification of the determinants of  $\theta_{jk}$ . Let  $L(j) := \min\{l \geq$

---

<sup>14</sup>Abbring and Van den Berg (2005) relate a similar argument for the case of simple random censoring to the dynamic selection problems studied by Ham and LaLonde (1996).

$0 : \Pr(S_l = j) > 0\}$  be the smallest number of transitions from  $S_0$  through which  $j$  is accessible. Then, we need data on at least  $L(j)+1$  spells to identify the determinants of  $\theta_{jk}$ . Take some  $u_0, u_1, \dots, u_{L(j)+1}$  such that  $\Pr(S_0 = u_0, \dots, S_{L(j)+1} = u_{L(j)+1}) > 0$ ,  $u_{L(j)} = j$ , and  $u_{L(j)+1} = k$ . We approach the identification of the determinants of  $\theta_{jk}$  by considering the identification of the determinants of

$$\Pr(S_0 = u_0, \dots, S_{L(j)+1} = u_{L(j)+1}; \Delta T_1 > t_1, \dots, \Delta T_{L(j)+1} > t_{L(j)+1} | X, V),$$

for  $(t_1, \dots, t_{L(j)+1}) \in \mathbb{R}_+^{L(j)+1}$ . In turn, these can be expressed in terms of the transition intensities corresponding to the origin-destination pairs in  $\Xi := \{(s, s') \in \mathcal{S}^2 : s \in \{u_0, \dots, u_{L(j)}\}, s' \in \mathcal{Z}(s)\}$ . Note that, along with the identification of  $\theta_{jk}$ 's determinants, we consider identification of the selection process into the state  $j$ . This selection process is modeled as a repeated competing-risks model, and its identification will again exploit results from the competing-risks literature.

We make the following assumptions.

**Assumption 5.**  $(V_p, p \in \Xi) \perp\!\!\!\perp X | S_0$ .

**Assumption 6.** For all  $s \in \mathcal{Z}(u_0)$ ,  $\mathbb{E}[V_{u_0s}] < \infty$ . For all  $s \in \mathcal{Z}(u_l)$ ,  $\mathbb{E}[V_{u_0u_1} \cdots V_{u_{l-1}u_l} V_{u_ls}] < \infty$ ,  $l = 1, \dots, L(j)$ .

These assumptions generalize Assumptions 1 and 2 to the case of multiple transitions, but for a single initial state  $u_0$ . Assumption 5 requires that the unobserved factors relevant to the chosen path from  $u_0$  to  $k$ , including those corresponding to the pairs in  $\mathcal{Z}(u_0)$ , are jointly independent of  $X$ , given  $S_0$ . Assumption 6 again facilitates inference on regressor effects at short durations. The higher moments appear in probabilistic expressions involving histories with multiple short spells. Together with Assumption 5, it allows us to derive the following result.

**Proposition 2.** *If Assumptions 5 and 6 are satisfied, then  $(\phi_p, p \in \Xi)$  is identified from the distribution of  $\{(\Delta T_l, S_l), l = 1, \dots, L(j) + 1\} | S_0, X$ .*

*Proof.* See Appendix. □

With a generalization of Assumption 3 to the case of multiple transitions, we can extend this result to identification of the full model. Let  $\Xi(u_l) := \{u_l\} \times \mathcal{Z}(u_l)$  be the set of all origin-destination pairs with origin  $u_l$ .

**Assumption 7.** For  $l = 0, \dots, L(j)$ , the set

$$\{(\Lambda_p(t_{i+1})\phi_p(x); p \in \Xi(u_i), i = 0, \dots, l); x \in \mathcal{X}, (t_1, \dots, t_l) \in \mathbb{R}_+^l, t_{l+1} = t^*\}$$

contains a nonempty open set in  $(0, \infty)^{\sum_{i=0}^l \#\Xi(u_i)}$ .

A sufficient condition for Assumption 7 is that the range  $\{(\phi_p(x), p \in \Xi), x \in \mathcal{X}\}$  of the regressor functions contains a nonempty open set in  $(0, \infty)^{\#\Xi}$ . If  $\phi_p(x) = \exp(x'\beta_p)$ , it would be sufficient that  $(\beta_p, p \in \Xi)$  has full column rank and  $\mathcal{X}$  contains a non-empty open set in  $\mathbb{R}^q$ . However, Assumption 7 is substantially weaker, as it allows us to substitute variation in the durations of previous spells for regressor variation.

We have the following result.

**Proposition 3.** *If Assumptions 4–7 are satisfied, then  $((\phi_p, \Lambda_p), p \in \Xi)$  and the joint distribution of  $(V_p, p \in \Xi)$  are identified from the distribution of  $\{(\Delta T_l, S_l), l = 1, \dots, L(j) + 1\} | S_0, X$ .*

*Proof.* See Appendix. □

The model for the first  $L(j)$  transitions is a repeated MPH competing-risks model and Proposition 3's proof iteratively applies an identification strategy similar to that of Abbring and Van den Berg (2003a) for the competing-risks model. In particular, identification of the determinants of the first transition from  $u_0$  is proven analogously to Proposition 1 (or, similarly, Abbring and Van den Berg, 2003a, Proposition 2). With this in hand, we can proceed to identification of the determinants of the

transition from  $u_1$ , exploiting knowledge of the determinants of the first transition, etcetera.

Proposition 3 establishes identifiability of the determinants of  $\theta_{jk}$  from event histories that include only a single spell in state  $j$ . If we have data on sufficiently long event histories, we may be able to observe multiple spells in state  $j$ . The literature on the identifiability of multi-spell duration models (Honoré, 1993; Abbring and Van den Berg, 2003a,b) suggests that many of our assumptions, including the proportional-hazards assumption, can be relaxed in this case. We will not further pursue this here.

## 4 Alternative Sampling Schemes

We have analyzed identifiability of the mixed semi-Markov event-history model from a random sample of censored event histories. In empirical practice, we often have to deal with alternative, more complex sampling schemes.

We distinguish two cases, inflow sampling and stock sampling (*e.g.* Lancaster, 1990, Chapter 8). With inflow sampling, we sample from the flow into a given subset of states during some time interval. Due to dynamic selection, the distribution of  $X, V$  in an inflow sample is typically not the same as its distribution in the population. With stock sampling, we sample from the stock in a given set of states at a certain moment in time. Again, the distribution of  $X, V$  in a stock sample is generally not the same as its population distribution. Moreover, conditional on  $X, V$  the distribution of the spells ongoing at the sampling date will not be the population distribution either in this case.

Various ways to model inflow and stock samples have been proposed in the literature (see *e.g.* Heckman and Singer, 1984a, 1986; Lancaster, 1990). In the case of inflow sampling, we could replace Assumption 5 by an *ad hoc* assumption on the distribution of the covariates in the inflow. We could make similar *ad hoc* assumptions

in the stock-sampling case, together with an assumption on the historical development of the inflow into the states from which we sample. One common assumption is that the inflow has been constant over time.

As Lancaster (1990, Section 8.4.2) points out, the common *ad hoc* assumptions on the distribution of the covariates are likely to be mutually inconsistent between the various sampling schemes. This suggests that we only make assumptions on the population, like Assumption 5, and derive the distributions of the various samples from the population model. In general, this is hard because the distributions of the inflow and stock samples depend on that of the full event history, including its initial conditions  $T_0, S_0$ .

An elegant solution is to assume that the samples are drawn from the event histories' long-run equilibrium distributions. This is only appropriate in carefully selected applications, and requires ergodicity of the semi-Markov model for the individual event histories. The resulting models for the inflow and stock samples are easy to derive and handle (Lancaster, 1990). In general, they involve dependent observable and unobservable covariates even under Assumption 5. However, this dependence is very tightly structured. We conjecture that this structure can be exploited to prove identifiability under conditions not unlike those for the case of sampling from the population. This is a topic for future research.

## 5 Concluding Remarks

This paper reviews the use of event-history models to simultaneously model dynamic selection into programs and the causal effects of the participation in such programs on event-history outcomes. A leading example is the analysis of the effects of training programs for the unemployed on their unemployment durations and subsequent job stability. We have provided novel identification results for a particular class of event-history models with a mixed semi-Markov structure. In doing so, we have highlighted

and exploited the central role of dependent competing-risks models.

We have focused on identification of causal and selection effects from “ideal”, large data sets, and have ignored sampling variation. Therefore, our results cannot be implemented directly in empirical practice. Instead, they explore the logical limits on what we can reasonably expect to learn about causal effects of dynamic programs from observational data. As such, they can guide the empirical analysis of causal program effects using appropriate event-history models.



# Appendix

## An Auxiliary Result

The proofs use a result for completely monotone functions. Completely monotone functions are frequently encountered in statistical duration analysis in the form of (derivatives of) Laplace transforms. They are formally defined by

**Definition 1.** Let  $\Omega$  be a nonempty open set in  $\mathbb{R}^n$ . A function  $f : \Omega \rightarrow \mathbb{R}$  is *absolutely monotone* if it is nonnegative and has nonnegative continuous partial derivatives of all orders.  $f$  is *completely monotone* if  $f \circ m$  is absolutely monotone, where  $m : x \in \{\omega \in \mathbb{R}^n : -\omega \in \Omega\} \mapsto -x$ .

Note that for  $n = 1$  this definition reduces to the familiar definition in Widder (1946). Abbring and Van den Berg (2003a, Proposition 1) state the following result.

**Proposition 4.** Let  $\Psi$  be a nonempty open connected set in  $\mathbb{R}^n$  and let  $f : \Psi \rightarrow \mathbb{R}$  and  $g : \Psi \rightarrow \mathbb{R}$  be completely monotone. If  $f$  and  $g$  agree on a nonempty open set in  $\Psi$ , then  $f = g$ .

Proposition 4's proof exploits two facts that are well-known for functions on  $\mathbb{R}$  and that are also true for functions on  $\mathbb{R}^n$ : (i) completely monotone functions are real analytic and (ii) real analytic functions are uniquely determined by their values on a nonempty open set.

## Proofs

For  $l \geq 0$ , let  $\mathcal{S}_0^l$  denote the support of  $(S_0, S_1, \dots, S_l)$ . As an extension of Tsiatis (1975), we can represent the “data” of Subsection 3.4's identification analysis, the distribution of  $\{(\Delta T_l, S_l), l = 1, \dots, L(j) + 1\} | (S_0, X)$ , by a collection  $\{Q_{\mathbf{s}}^{L(j)+1}(\cdot | x); \mathbf{s} \in \mathcal{S}_0^{L(j)+1}, x \in \mathcal{X}\}$  such that  $Q_{\mathbf{s}}^{L(j)+1}(\mathbf{t} | X)$  equals

$$\Pr(\Delta T_1 > t_1, S_1 = s_1; \dots; \Delta T_{L(j)+1} > t_{L(j)+1}, S_{L(j)+1} = s_{L(j)+1} | X, S_0 = s_0)$$

almost surely, for all  $\mathbf{s} := (s_0, \dots, s_{L(j)+1}) \in \mathcal{S}_0^{L(j)+1}$  and  $\mathbf{t} := (t_1, \dots, t_{L(j)+1}) \in \mathbb{R}_+^{L(j)+1}$ .

From these data, we can derive  $Q_{\mathbf{s}_0^l}^l$ , defined analogously to  $Q_{\mathbf{s}}^{L(j)+1}$ , and  $R_{\mathbf{s}_0^{l-1}}^l := \sum_{s \in \mathcal{Z}(s_{l-1})} Q_{\mathbf{s}_0^{l-1}s}^l$ , for  $l = 1, 2, \dots, L(j)+1$ ,  $\mathbf{s}_0^{l-1} := (s_0, \dots, s_{l-1}) \in \mathcal{S}_0^{l-1}$ , and  $\mathbf{s}_0^l \in \mathcal{S}_0^l$ . Note that  $R_{\mathbf{s}_0^{l-1}}^l(\mathbf{t}_1^l | X)$  equals

$$\Pr(\Delta T_1 > t_1, S_1 = s_1; \dots; \Delta T_{l-1} > t_{l-1}, S_{l-1} = s_{l-1}; T_l > t_l | X, S_0 = s_0)$$

almost surely, for all  $\mathbf{s}_0^{l-1} \in \mathcal{S}_0^{l-1}$ ,  $\mathbf{t}_1^l := (t_1, \dots, t_l) \in \mathbb{R}_+^l$ , and  $l = 1, 2, \dots, L(j)+1$ .

*Proof of Proposition 2.* The proof proceeds iteratively. First, consider identification of  $\phi_p$  for  $p \in \Xi(u_0)$ . Pick an arbitrary  $x \in \mathcal{X}$ . Note that  $Q_p^1(\cdot | x)$  and  $Q_p^1(\cdot | x^*)$  are differentiable almost everywhere, and that

$$\phi_p(x) = \lim_{t \downarrow 0} \frac{\partial Q_p^1(t | x) / \partial t}{\partial Q_p^1(t | x^*) / \partial t}.$$

because of Assumptions 5 and 6 and the normalization  $\phi_p(x^*) = 1$ . Because  $x$  is arbitrary, this identifies  $(\phi_p, p \in \Xi(u_0))$ .

Next, iterate the following argument for  $l = 1, \dots, L(j)$ . Suppose that  $\phi_{u_0 u_1}, \dots, \phi_{u_{l-1} u_l}$  are identified and consider identification of  $\phi_p$  for  $p \in \Xi(u_l)$ . Pick an arbitrary  $x \in \mathcal{X}$ . Note that  $Q_{\mathbf{s}}^{l+1}(\cdot | x)$  and  $Q_{\mathbf{s}}^{l+1}(\cdot | x^*)$  are differentiable almost everywhere, and that

$$\phi_{u_0 u_1}(x) \cdots \phi_{u_{l-1} u_l}(x) \phi_{u_l s}(x) = \lim_{\mathbf{t} \downarrow 0} \frac{\partial^{l+1} Q_{\mathbf{s}}^{l+1}(\mathbf{t} | x) / \partial t_1 \cdots \partial t_{l+1}}{\partial^{l+1} Q_{\mathbf{s}}^{l+1}(\mathbf{t} | x^*) / \partial t_1 \cdots \partial t_{l+1}},$$

with  $\mathbf{s} = (u_0, \dots, u_l, s)$ ,  $s \in \mathcal{Z}(u_l)$ , and  $\mathbf{t} = (t_1, \dots, t_{l+1})$ . Here, we have used Assumptions 5 and 6 and the normalizations  $\phi_{u_0 u_1}(x^*) = \cdots = \phi_{u_{l-1} u_l}(x^*) = \phi_{u_l s}(x^*) = 1$ . Because  $x$  is arbitrary, this identifies  $(\phi_p, p \in \Xi(u_l))$ .  $\square$

*Proof of Proposition 3.*  $(\phi_p, p \in \Xi)$  is identified by Proposition 2. The remainder of the proof again proceeds iteratively.

First, Assumptions 4 and 5 and the normalizations ( $\Lambda_p(t^*) = 1, p \in \Xi(u_0)$ ) imply that

$$R_{u_0}^1(t^*|x) = \mathcal{L}_{\Xi(u_0)}(\phi_p(x), p \in \Xi(u_0)),$$

where  $\mathcal{L}_{\Xi(u_0)}$  is the Laplace transform of the joint distribution of  $(V_p, p \in \Xi(u_0))$ . Note that  $R_{u_0}^1(t^*|\cdot)$  and  $(\phi_p, p \in \Xi(u_0))$  are identified at this point. So, by Assumption 7 we can trace out  $\mathcal{L}_{\Xi(u_0)}$  on a nonempty open set in  $(0, \infty)^{\#\Xi(u_0)}$ . Because  $\mathcal{L}_{\Xi(u_0)}$  is completely monotone, this identifies  $\mathcal{L}_{\Xi(u_0)}$  by Proposition 4. By implication,  $(D_p \mathcal{L}_{\Xi(u_0)}, p \in \Xi(u_0))$  is identified, with  $D_p \mathcal{L}_{\Xi(u_0)}$  the partial derivative of  $\mathcal{L}_{\Xi(u_0)}$  with respect to the argument corresponding to  $V_p, p \in \Xi(u_0)$ .

Pick an arbitrary  $x$ . For almost all  $t \in (0, \infty)$

$$\Lambda'_p(t) = \frac{\partial Q_p^1(t|x)/\partial t}{\phi_p(x) D_p \mathcal{L}_{\Xi(u_0)}(\Lambda_{p'}(t) \phi_{p'}(x), p' \in \Xi(u_0))}, \quad p \in \Xi(u_0),$$

by Assumptions 4 and 5. These  $\#\Xi(u_0)$  equations form a system of differential equations in  $(\Lambda'_p, p \in \Xi(u_0))$ ,  $(\Lambda_p, p \in \Xi(u_0))$ , and  $t$ , with initial conditions  $(\Lambda_p(t^*) = 1, p \in \Xi(u_0))$ , in the sense of Carathéodory (1918). Analogously to Abbring and Van den Berg (2003a, Proposition 2), this system can be shown to have a unique solution  $(\Lambda_p, p \in \Xi(u_0))$  in terms of  $(Q_p^1, D_p \mathcal{L}_{\Xi(u_0)}, \phi_p(x); p \in \Xi(u_0))$ . Because the latter have been identified at this point, this establishes identification of  $(\Lambda_p, p \in \Xi(u_0))$ .

Second, iterate the following argument for  $l = 1, \dots, L(j)$ . Suppose that  $(\Lambda_p, p \in \Xi(\mathbf{u}_0^{l-1}))$  is identified, with  $\mathbf{u}_0^{l-1} := (u_0, \dots, u_{l-1})$  and  $\Xi(\mathbf{u}_0^{l-1}) := \bigcup_{i=0}^{l-1} \Xi(u_i)$ . By Assumptions 4 and 5, the  $l$ -th partial derivative  $D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}$  of  $\mathcal{L}_{\Xi(\mathbf{u}_0^l)}$  with respect to the arguments corresponding to  $V_{u_0 u_1}, \dots, V_{u_{l-1} u_l}$  satisfies

$$D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}(\Lambda_p(t_{i+1}) \phi_p(x); p \in \Xi(u_i), i = 0, \dots, l) = \frac{\partial R_{\mathbf{u}_0^l}^{l+1}(\mathbf{t}_1^{l+1}|x)/\partial t_1 \cdots \partial t_l}{\prod_{i=1}^l \Lambda_{u_{i-1} u_i}(t_i) \phi_{u_{i-1} u_i}(x)}.$$

for almost all  $\mathbf{t}_1^l \in \mathbb{R}_+^l$ , all  $t_{l+1} \in \mathbb{R}_+$ , and all  $x \in \mathcal{X}$ . Because the right-hand side is known at this point, this identifies

$$D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}(\Lambda_p(t_{i+1})\phi_p(x); p \in \Xi(u_i), i = 0, \dots, l)$$

for all  $\mathbf{t}_1^{l+1} \in \mathbb{R}_+^{l+1}$  and  $x \in \mathcal{X}$ . Moreover, with the normalizations  $(\Lambda_p(t^*) = 1, p \in \Xi(u_l))$ ,  $(\Lambda_p(t_{i+1})\phi_p(x); p \in \Xi(u_i), i = 0, \dots, l)$  is identified for all  $\mathbf{t}_1^l \in \mathbb{R}_+^l$ ,  $t_{l+1} = t^*$ , and  $x \in \mathcal{X}$  at this point. So, by Assumption 7, we can trace out  $D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}$  on a nonempty open set in  $(0, \infty)^{\#\Xi(\mathbf{u}_0^l)}$ . Because  $(-1)^l D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}$  is completely monotone, this identifies  $D_{\mathbf{u}_0^l} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}$  by Proposition 4. By implication,  $(D_{\mathbf{u}_0^l s} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}, s \in \mathcal{Z}(u_l))$  is identified.

Pick an arbitrary  $x$ . Pick  $t_i$  such that  $\Lambda_{u_{i-1}u_i}$  is differentiable at  $t_i$ ,  $i = 1, \dots, l$ . For almost all  $t_{l+1} \in (0, \infty)$

$$\begin{aligned} \Lambda'_{u_l s}(t_{l+1}) &= \left[ \phi_{u_l s}(x) \prod_{i=1}^l \Lambda'_{u_{i-1}u_i}(t_i) \phi_{u_{i-1}u_i}(x) \right]^{-1} \times \\ &\quad \frac{\partial Q_{\mathbf{u}_0^l s}^{l+1}(\mathbf{t}_1^{l+1}|x) / \partial t_1 \cdots \partial t_{l+1}}{D_{\mathbf{u}_0^l s} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}(\Lambda_p(t_{i+1})\phi_p(x); p \in \Xi(u_i), i = 0, \dots, l)}, \quad s \in \mathcal{Z}(u_l), \end{aligned} \quad (2)$$

by Assumptions 4 and 5. These  $\#\Xi(u_l)$  equations again form a system of differential equations in the sense of Carathéodory (1918), now in  $(\Lambda'_p, p \in \Xi(u_l))$ ,  $(\Lambda_p, p \in \Xi(u_l))$ , and  $t_{l+1}$ , with initial conditions  $(\Lambda_p(t^*) = 1, p \in \Xi(u_l))$ . Standard theory can again be applied to show that this system has a unique solution  $(\Lambda_p, p \in \Xi(u_l))$  in terms of

$$\left( Q_{\mathbf{u}_0^l s}^{l+1}, D_{\mathbf{u}_0^l s} \mathcal{L}_{\Xi(\mathbf{u}_0^l)}, \phi_{u_l s}(x) \prod_{i=1}^l \Lambda'_{u_{i-1}u_i}(t_i) \phi_{u_{i-1}u_i}(x) ; s \in \mathcal{Z}(u_l) \right).$$

Because the latter is identified at this point, this establishes identification of  $(\Lambda_p, p \in \Xi(u_l))$ .

Finally note that  $\mathcal{L}_{\Xi}$  is identified by integrating  $D_{\mathbf{u}_0^{L(j)}} \mathcal{L}_{\Xi}$ . In turn,  $\mathcal{L}_{\Xi}$  identifies

the joint distribution of  $(V_p, p \in \Xi)$  by the uniqueness of the multivariate Laplace transform. □

## References

- Abbring, J. H. (2002). Stayers versus defecting movers: A note on the identification of defective duration models. *Economics Letters* 74, 327–331.
- Abbring, J. H. (2003). Dynamic econometric program evaluation. Working paper, Free University, Amsterdam. Paper prepared for the H. Theil Memorial Conference, Amsterdam, 16-18 August 2002.
- Abbring, J. H. and J. J. Heckman (2006). Dynamic policy analysis. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data* (third ed.). Dordrecht: Kluwer. Forthcoming.
- Abbring, J. H. and G. J. Van den Berg (2003a). The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society Series B* 65, 701–710.
- Abbring, J. H. and G. J. Van den Berg (2003b). The non-parametric identification of treatment effects in duration models. *Econometrica* 71, 1491–1517.
- Abbring, J. H. and G. J. Van den Berg (2004). Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models. *Empirical Economics* 29, 5–20.
- Abbring, J. H. and G. J. Van den Berg (2005). Social experiments and instrumental variables with duration outcomes. Discussion Paper 2005-047/3, Tinbergen Institute, Amsterdam.
- Abbring, J. H., G. J. Van den Berg, and J. C. Van Ours (2005). The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *Economic Journal* 115, 602–630.
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

- Bonnal, L., D. Fougère, and A. Sérandon (1997). Evaluating the impact of French employment policies on individual labour market histories. *Review of Economic Studies* 64, 683–713.
- Carathéodory, C. (1918). *Vorlesungen über Reelle Funktionen*. Leipzig: Teubner.
- Card, D. and D. Sullivan (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica* 56, 497–530.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B* 34, 187–202.
- Eberwein, C., J. C. Ham, and R. J. Lalonde (1997). The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data. *Review of Economic Studies* 64, 655–682.
- Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* 64, 403–409.
- Gritz, R. M. (1993). The impact of training on the frequency and duration of employment. *Journal of Econometrics* 57, 21–51.
- Ham, J. C. and R. J. LaLonde (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica* 64, 175–205.
- Heckman, J. J. (1981). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in Labor Markets*, National Bureau of Economic Research, pp. 91–139. University of Chicago Press.
- Heckman, J. J. and G. J. Borjas (1980, August). Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model

- of heterogeneity and state dependence. *Economica* 47(187), 247–283. Special Issue on Unemployment.
- Heckman, J. J. and B. E. Honoré (1989, June). The identifiability of the competing risks model. *Biometrika* 76(2), 325–330.
- Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 31, pp. 1865–2097. New York: North-Holland.
- Heckman, J. J. and B. S. Singer (1984a, January-February). Econometric duration analysis. *Journal of Econometrics* 24(1-2), 63–132.
- Heckman, J. J. and B. S. Singer (1984b, April). The identifiability of the proportional hazard model. *Review of Economics Studies* 51(2), 231–243.
- Heckman, J. J. and B. S. Singer (1986, June). Econometric analysis of longitudinal data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 3, Chapter 29, pp. 1690–1763. North-Holland.
- Heckman, J. J. and C. Taber (1994). Econometric mixture models and more general models for unobservables in duration analysis. *Statistical Methods in Medical Research* 3(3), 279–299.
- Heckman, J. J. and E. J. Vytlačil (2006). Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*. Amsterdam: Elsevier. forthcoming.
- Honoré, B. E. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies* 60, 241–246.



- Keiding, N. (1999). Event history analysis and inference from observational epidemiology. *Statistics in Medicine* 18, 2353–2363.
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* 47, 939–956.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica* 58, 757–782.
- Ridder, G. (1986). An event history approach to the evaluation of training, recruitment and employment programmes. *Journal of Applied Econometrics* 1, 109–126.
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167–182.
- Ridder, G. and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica* 71, 1579–1589.
- Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 72, 20–22.
- Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple durations. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume V. North Holland.
- Van den Berg, G. J., A. Holm, and J. C. Van Ours (2002). Do stepping-stone jobs exist? Early career paths in the medical profession. *Journal of Population Economics* 15, 647–665.
- Van den Berg, G. J., B. Van der Klaauw, and J. C. Van Ours (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics* 22, 211–241.

- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16, 439–454.
- Widder, D. V. (1946). *The Laplace Transform*. Princeton: Princeton University Press.